1257692

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

December 07, 2004

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *60/513,237*
FILING DATE: *October 21, 2003*
RELATED PCT APPLICATION NUMBER: *PCT/US04/34913*

Certified by

Jon W Dudas

Acting Under Secretary of Commerce
for Intellectual Property
and Acting Director of the U.S.
Patent and Trademark Office

SUBSTITUTE PTO/SB/16 (5-03)

17698 U.S. PTO 102103

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR §1.53(c).

09/746,[illegible] US PTO 60/513237 102103

INVENTOR(S)

Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Mark J.	Burk	San Diego, CA
Michael	Levin	San Diego, CA
Zoulin	Zhu	San Diego, CA
Jennifer Ann	Chaplin	San Diego, CA
Karen	Kustedjo	La Jolla, CA
Zilin	Huang	San Diego, CA

Additional inventors are being named on the 0 separately numbered sheets attached hereto

TITLE OF THE INVENTION (500 characters max)

METHODS FOR MAKING SIMVASTATIN AND INTERMEDIATES

CORRESPONDENCE ADDRESS

Direct all correspondence to:

[X] Customer Number: 20985

OR

[ ] Firm *or* Individual Name					
Address					
Address					
City		State		ZIP	
Country	United States	Telephone		Fax	

ENCLOSED APPLICATION PARTS *(check all that apply)*

[X] Specification *Number of Pages* 23

[ ] Drawing(s) *Number of Sheets*

[ ] Application Data Sheet. See 37 CFR 1.76.

[ ] CD(s), Number

[X] Other (specify)

2 pgs. claims; 1 pg. Abstract; Appendix A - 23 pgs.; Appendix B – 13 pgs; Sequence Listing – 11 pgs.

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT

[X] Applicant Claims small entity status. See 37 CFR 1.27.

[X] A check or money order is enclosed to cover the filing fees.

[ ] The Director is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number: 06-1050

[ ] Payment by credit card. Form PTO-2038 is attached.

FILING FEE AMOUNT ($) $80

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

[X] No.

[ ] Yes, the name of the U.S. Government agency and the Government contract number are:

Respectfully submitted,

Signature [signature]

Date October 21, 2003

Typed Name Gregory P. Einhorn, Reg. No. 38,440

Telephone No. (858) 678-5070

Docket No. 09010-983P01

10339753.doc

CERTIFICATE OF MAILING BY EXPRESS MAIL

Express Mail Label No. EV399292089US

Date of Deposit October 21, 2003

PROVISIONAL APPLICATION FOR UNITED STATES PATENT

under 37 CFR §1.53(c)

for

METHODS FOR MAKING SIMVASTATIN AND INTERMEDIATES

Inventors:. **Mark Burk**
Michael Levin
Zoulin Zhu
Jennifer Chaplin
Karen Kustedjo
Zilin Huang
Brian Morgan

Assignee: Diversa Corporation
4955 Directors Place
San Diego, California 92121 U.S.A

Fish & Richardson P.C.
12390 El Camino Real
San Diego, California 92130-2081
Tel.: (858) 678-5070
Fax: (858) 678-5099

ATTORNEY DOCKET:
09010-983P01

DATE OF DEPOSIT: Oct. 21, 2003

EXPRESS MAIL NO.: EV399242089 US

METHODS FOR MAKING SIMVASTATIN AND INTERMEDIATES

TECHNICAL FIELD

This invention generally relates to the field of synthetic organic and medicinal chemistry. In one aspect, the invention provides synthetic chemical and chemoenzymatic methods of producing simvastatin and various intermediates. In one aspect, enzymes such as hydrolases are used in the methods of the invention.

BACKGROUND

Simvastatin is a potent antihypercholesterolemic agent that is presently marketed under the name ZOCOR®. Simvastatin, Mevastatin, Lovastatin and Pravastatin are hexahydronaphthalene derivatives used as inhibitors of the enzyme HMG-CoA reductase, the rate-controlling enzyme in the biosynthetic pathway for formation of cholesterol in the human body.

Mevastatin, Lovastatin and Pravastatin are natural fermentation products which possess a 2-methylbutyrate side chain at C-8 of their hexahydronaphthalene ring system. Compounds possessing a C-8 2,2-dimethylbutyrate side chain, including Simvastatin, can be better inhibitors of HMG-CoA reductase than their 2-methylbutyrate counterparts. Thus 2,2-dimethylbutyrate derivatives may have greater promise for the treatment of atherosclerosis, hyperlipemia, familial hypercholesterolemia and similar disorders. However, these derivatives, including Simvastatin, are not naturally occurring and have to be produced synthetically. As a result, the introduction on the market of the more potent HMG-CoA reductase inhibitor Simvastatin has prompted the need for efficient, high yielding processes for manufacturing it.

SUMMARY

The invention provides methods for the preparation of Simvastatin, including at least one method as set forth in Appendix A or Appendix B. In one aspect, diol lactone is regioselectively acylated at the 8-position using a derivative of dimethylbutyric acid and a Lewis acid catalyst.

In alternative aspects of any of the methods of the invention, at least one step is performed in a reaction vessel. In alternative aspects of any of the methods of the invention, at least one step is performed in a cell extract. In alternative aspects of any of the methods of the invention, at least one step is performed in a whole cell. The cell can be of any source, e.g., a plant cell, a bacterial cell, a fungal cell, a mammalian cell or a yeast cell.

In one aspect of any of the methods of the invention, an ammonium salt of simvastatin is formed.

In one aspect, the methods further comprise re-crystallization of the simvastatin. In one aspect, the methods comprise relactonization to provide simvastatin with a desired purity.

In one aspect of any of the methods of the invention, at least one enzymatic reaction is carried out by a hydrolase encoded by a nucleic acid having at least 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or complete (100%) sequence identity to SEQ ID NO:1, or enzymatically active fragments thereof. In one aspect of any of the methods of the invention, at least one enzymatic reaction is carried out by a hydrolase encoded by a nucleic acid having at least 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or complete (100%) sequence identity to SEQ ID NO:3, or enzymatically active fragments thereof.. In one aspect of any of the methods of the invention, at least one enzymatic reaction is carried out by a hydrolase encoded by a nucleic acid having at least 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or complete (100%) sequence identity to SEQ ID NO:5, or enzymatically active fragments thereof..

In one aspect of any of the methods of the invention, at least one enzymatic reaction is carried out by a hydrolase having a sequence at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%,

70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or complete (100%) sequence identity to SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6, or enzymatically active fragments thereof..

The details of one or more embodiments of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

All publications, patents, patent applications, GenBank sequences and ATCC deposits, cited herein are hereby expressly incorporated by reference for all purposes.

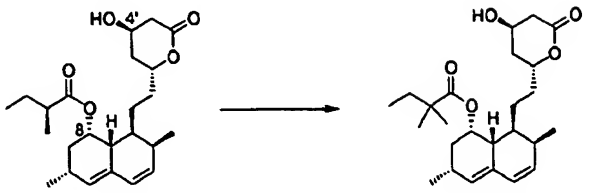
DETAILED DESCRIPTION

The present invention provides novel synthetic chemical and biochemical processes for the production of Simvastatin and its intermediates. These methods can be efficient and cost-effective.

In various aspects of the invention, the methods catalyze reactions biocatalytically using various enzymes, including hydrolases, e.g., acylases and esterases. In one aspect, the invention provides methods for the enzymatic hydrolysis of lovastatin to lovastatin acid using hydrolases. In one aspect, the invention provides methods for the enzymatic acylation of diol lactone to an acyl lactone using hydrolases. In one aspect, the invention provides methods for the enzymatic acylation of an acyl lactone to an acyl simvastatin using hydrolases. In one aspect, the invention provides methods for hydrolyzing a lactone ring using hydrolases.

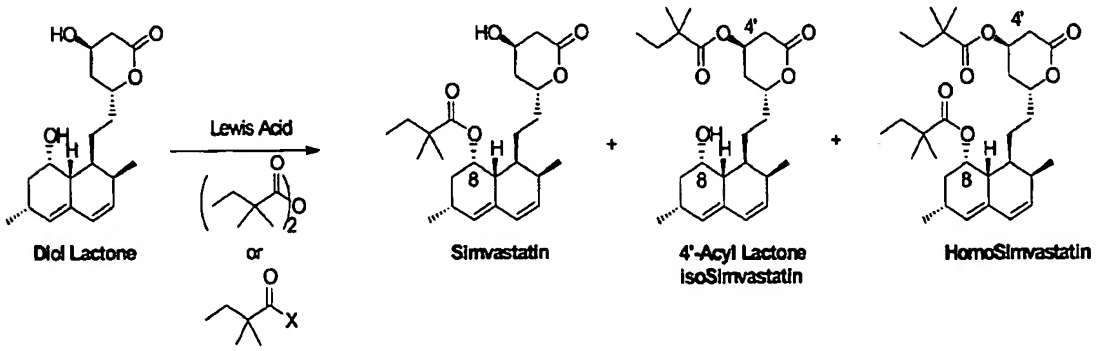
The invention includes methods for producing simvastatin and various intermediates via *in vitro* or *in vivo* techniques, e.g., whole cells protocols, such as fermentation or other biocatalytic processes.

In one aspect, the invention provides processes comprising a short, convenient route for the conversion of lovastatin into simvastatin, including:



In one aspect, diol lactone made from lovastatin via hydrolysis is regioselectively acylated at the 8-position using a derivative of dimethylbutyric acid and a Lewis acid catalyst. Diol lactone can be made from lovastatin using chemoenzymatic processes described herein.

In one aspect, the invention provides a process comprising:

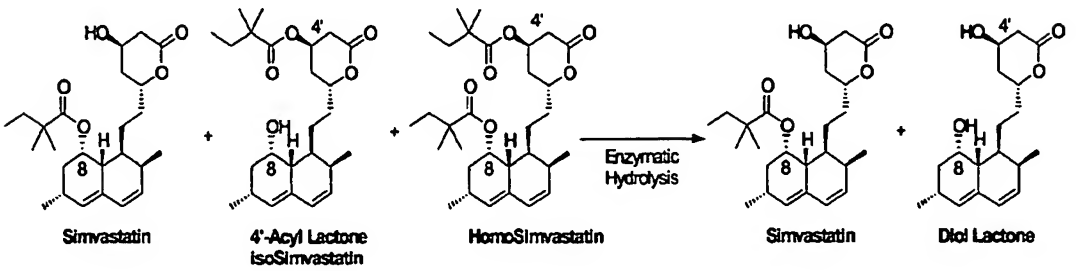


The inventors have found that the treatment of diol lactone with a carboxylic acid derivative in the presence of a Lewis acid catalyst results in predominant acylation at the 8-position. When excess vinyl acetate is used in the presence of a metal triflate, the 8-acetyl derivate is formed almost exclusively at low conversion. Results to date show that the treatment of diol lactone with a combination of dimethylbutyric anhydride, and $Bi(OTf)_3$ or $Cu(OTf)_2$ in dichloromethane at room temperature results in a rapid reaction in which the simvastatin: 4'-acyl lactone ratio is >4:1.

In one aspect, the isolation and purification of simvastatin is by crystallization. In one aspect, the invention provides methods for screening Lewis acid catalysts and/or acylation agents to provide alternative reaction conditions to maximize the yield of simvastatin and minimize the side products. Maximizing the yield of simvastatin and minimizing the side products helps in crystallization protocols. Use of crystallization to

isolate/ purify simvastatin results in an exemplary 2-step process from lovastatin to simvastatin.

In one aspect, the invention provides a process comprising:

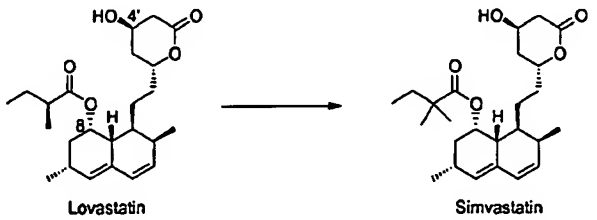


In one aspect, if isosimvastatin and homosimvastatin cannot be reduced to levels that can be purged by crystallization, a final enzymatic hydrolysis step is employed to facilitate the recovery of product. In one aspect, the treatment of mixtures of simvastatin, isosimvastatin and homosimvastatin with an esterase (e.g., enzyme having a sequence as set forth in SEQ ID NO:2, encoded by SEQ ID NO:1), results in the regioselective hydrolysis of the acyl group at the 4'-position, resulting in a mixture of simvastatin and diol lactone. In one aspect, the simvastatin is separated by crystallization.

Alternatively, the use of excess anhydride can be used to push the reaction towards the formation of simvastatin and homosimvastatin. This can minimize the amount of isosimvastatin. Enzymatic hydrolysis of such mixtures results in the formation and ready isolation of simvastatin.

In one aspect of the preparation of simvastatin by regioselective acylation of diol lactone in the presence of Lewis acids, Diol lactone was treated with dimethylbutyric anhydride (0.5 eq) in dichloromethane at room temperature (RT) in the presence of 5 mol% $Cu(OTf)_2$ as catalyst. HPLC analysis indicated 50% conversion of diol lactone within 10 minutes. The ratio of simvastatin (acylation at the 8-position) to isosimvastatin (acylation at the 4-position), was 4:1, with ~4% homosimvastatin being formed.

In one aspect, the invention provides a process comprising:



and alternative aspects, at least one, several or all, of the following steps:

Step 1: Enzymatic hydrolysis of lovastatin, lovastatin acid or a salt of lovastatin acid to form the triol acid using a hydrolase enzyme, e.g., an enzyme of the invention, e.g., SEQ ID NO: __.

Step 2: Heating the triol acid or stirring in the presence of acid to form the diol lactone

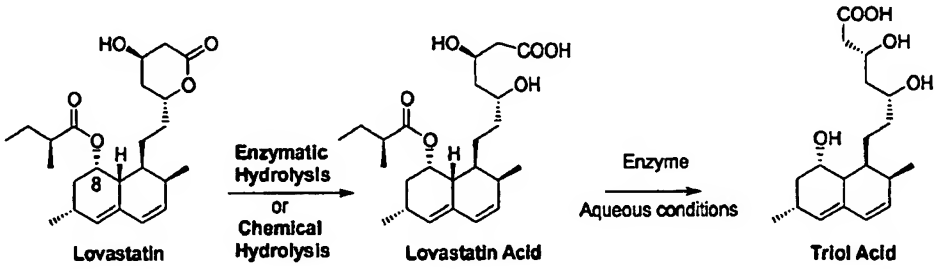
Step 3: Protection of the 4'-OH on the lactone ring by regioselective acylation, either chemically or by using a proprietary or commercially available hydrolase

Step 4: Acylation of the hydroxyl at the 8-position; can be carried out chemically, or enzymatically using a proprietary hydrolase

Step 5: Selective removal of the acyl protecting group at the 4' position, either chemically or enzymatically, yields simvastatin. If necessary, formation of the ammonium salt of simvastatin, and recrystallization of simvastatin, followed by re-lactonization, provides simvastatin with the desired purity.

In one aspect, referring to step 1, as described above, the invention provides a process comprising:

Step 1:



Complete, or substantially complete (in alternative aspects, >99%, >98%, >97% or >96%) removal of the methylbutyrate sidechain may be essential for a process because of the difficulty in separating lovastatin and simvastatin, and the low allowable levels of lovastatin in simvastatin API. Reported procedures for the hydrolysis of lovastatin require the use of high temperatures and long reaction times for complete reaction.

In one aspect, Lovastatin is hydrolyzed under mild conditions using a hydrolase enzyme (e.g., enzyme having a sequence as set forth in SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6, encoded by SEQ ID NO:1, SEQ ID NO:3 or SEQ ID NO:5,

respectively). This results in hydrolysis of the lactone ring and complete removal of the side-chain in the 8-position. The enzymes having a sequence as set forth in SEQ ID NO:1, SEQ ID NO:3 and SEQ ID NO:5 have been demonstrated to be particularly effective for the enzymatic hydrolysis of the methylbutyrate sidechain: SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6. The enzyme having a sequence as set forth in SEQ ID NO:2 has been subcloned and expressed in a *Pseudomonas* host and produced on a 10 liter (L) fermentation scale.

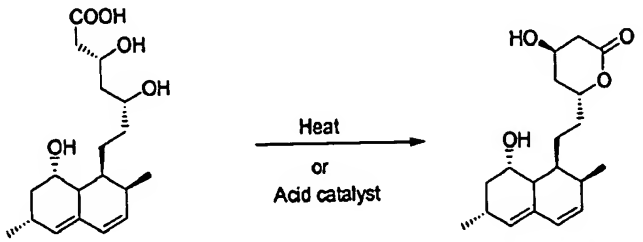
Lovastatin can show poor solubility under the aqueous conditions necessary for enzymatic activity. Thus, in one alternative aspect, a suspension of lovastatin in water is raised to pH >12 to effect a rapid hydrolysis of the lactone ring. This results in the *in-situ* formation of the more soluble lovastatin acid salt. In one aspect, the pH of the reaction mixture is then readjusted downward to a range suitable for the enzymatic reaction; and the enzyme is added.

The enzymatic hydrolysis conditions may also be applied to mixtures of lovastatin and lovastatin acid extracted directly from fermentation broth. Alternatively, the enzyme may be added to the fermentation broth and the triol acid isolated directly.

In one aspect, after hydrolysis, the reaction mixture is carefully acidified. The triol acid can be isolated by extraction and/or filtration and used directly in the next step. Alternatively, it the triol acid is isolated as a solid after a suitable crystallization/precipitation step.

In one aspect, referring to step 2, as described above, the invention provides a process comprising:

Step 2:

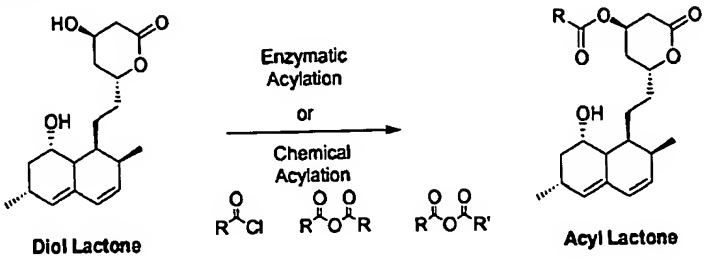


Triol Acid **Diol Lactone**

In one aspect, the triol acid is re-lactonized by heating in a suitable solvent and driving the equilibrium to the lactone form by removal of water by conventional means. Alternatively, stirring in the presence of a suitable acid will effect closure of the lactone ring.

In one aspect, referring to step 3, as described above, the invention provides a process comprising:

Step 3:



Regioselective acylation of the hydroxyl group in the 4'-position may be carried out chemically using a carboxylic acid derivative (e.g., acid chloride, symmetric or unsymmetric anhydride etc.), or enzymatically using an enzyme with the desired activity and selectivity, e.g., a hydrolase, such as an esterase. In one aspect, hydrolases (e.g., esterases) are used to acylate diol lactones. The nature of the acyl group can be varied to impart suitable properties, e.g., acetate for ease of removal, benzoate for enhanced crystallinity, formate for enhanced water solubility.

In alternative aspects of the exemplified methods described herein, including the reactions and reagents as illustrated in Steps 3 (supra), 4 and 5 (infra), "R" can be:

(i) - H, a formyl derivative;

(ii) a C1-n alkyl, both straight chain and branched;

(iii) substituted alkyl groups, e.g., chloroacetyl, trichloroacetyl, trifluoroacetyl, methoxyacetyl, phenylacetyl, 4-oxopentyl (levulinate);

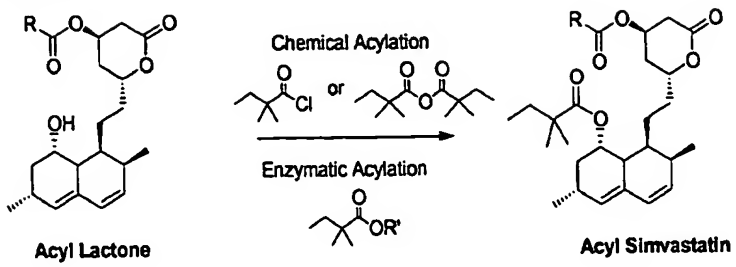
(iv) phenyl and substituted phenyl: e.g., phenyl, p-nitrophenyl;

(v) an R'O- group, forming a carbonate protecting group, exemplified but not limited to: tBuOCO, PhOCO, $PhCH_2OCO$.

In one aspect, an enzyme with enhanced reactivity on long-chain alkyl esters is used when R is a long-chain alkyl group. Solubility may a problem when R is a long-chain alkyl group. In one aspect, R is an acetate, which can be advantageous due to (i) ease of installation, (ii) good enzyme activity for hydrolysis, (iii) solubility, (iv) cost of reagents.

In one aspect, referring to step 4, as described above, the invention provides a process comprising:

Step 4:



In one aspect, a combination of a dimethylbutyric acid derivative with a suitable acylation catalyst (by chemical acylation or enzymatic acylation) is used to install the desired simvastatin side-chain. The combination of dimethylbutyric anhydride/Lewis acid (e.g., $Bi(triflate)_3$, $Cu(triflate)_2$), results in rapid reaction at room temperature (RT).

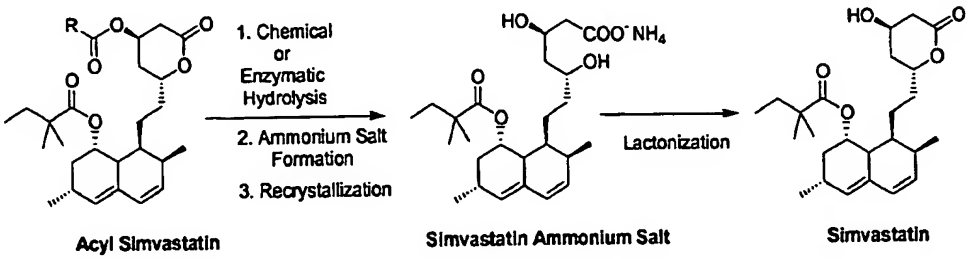
In one aspect, the invention provides methods for screening suitable Lewis acids and reaction conditions, including temperature, solvents etc. Optimum conditions for this acylation for alternative protocols or reagents can be determined using routing screening methods.

In one aspect, enzyme catalyzed acylation of the acyl lactone is used to install the dimethylbutyrate group at the 8-position under very mild conditions (for example, in one aspect, at RT, e.g., about 40°C, using organic solvent), without formation of side products.

The invention provides methods for screening for alternative enzymes that have the desired activity in the methods of the invention. Enzymes can be screened for their effectiveness in various protocols of the invention using routine methods.

In one aspect, referring to step 5, as described above, the invention provides a process comprising:

Step 5:



In one aspect, the final steps require the selective removal of the acyl group at the 4'-position. The acyl group at the 4'-position can be highly susceptible to base-catalyzed

elimination, even under only slightly basic conditions. Consequently, the enzymatic hydrolysis has been the most convenient method for regioselective removal of this acyl group. It has been demonstrated that the same enzyme that hydrolyzes lovastatin (SEQ ID NO:2 (encoded by SEQ ID NO:1), in step 1, above) is also an effective catalyst for the selective hydrolysis of acyl groups at the lactone 4'-position. When carried out at pH 7, this enzymatic hydrolysis yields simvastatin with the lactone ring substantially intact.

General Methods

The present invention provides novel biochemical processes for the production of simvastatin and various intermediates. The skilled artisan will recognize that the starting and intermediate compounds used in the methods of the invention can be synthesized using a variety of procedures and methodologies, which are well described in the scientific and patent literature., *e.g.*, Organic Syntheses Collective Volumes, Gilman *et al.* (Eds) John Wiley & Sons, Inc., NY; Venuti (1989) *Pharm Res.* 6:867-873. The invention can be practiced in conjunction with any method or protocol known in the art, which are well described in the scientific and patent literature.

The discussion of the general methods given herein is intended for illustrative purposes only. Other alternative methods and embodiments will be apparent to those of skill in the art upon review of this disclosure.

Enzymes used in the methods of the invention can be produced by any synthetic or recombinant method, or, they may be isolated from a natural source, or, a combination thereof. Nucleic acids encoding enzymes used to practice the methods of the invention, whether RNA, cDNA, genomic DNA, vectors, viruses or hybrids thereof, may be isolated from a variety of sources, genetically engineered, amplified, and/or expressed/ generated recombinantly. Recombinant polypeptides generated from these nucleic acids can be individually isolated or cloned and tested for a desired activity. Any recombinant expression system can be used, including bacterial, mammalian, yeast, insect or plant cell expression systems. Nucleic acids used to practice the methods of the invention can be generated using amplification methods, which are also well known in the art, and include, e.g., polymerase chain reaction, PCR (see, e.g., PCR PROTOCOLS, A GUIDE TO METHODS AND APPLICATIONS, ed. Innis, Academic Press, N.Y. (1990) and PCR STRATEGIES (1995), ed. Innis, Academic Press, Inc., N.Y., ligase chain reaction (LCR)

(see, e.g., Wu (1989) Genomics 4:560; Landegren (1988) Science 241:1077; Barringer (1990) Gene 89:117); transcription amplification (see, e.g., Kwoh (1989) Proc. Natl. Acad. Sci. USA 86:1173); and, self-sustained sequence replication (see, e.g., Guatelli (1990) Proc. Natl. Acad. Sci. USA 87:1874); Q Beta replicase amplification (see, e.g., Smith (1997) J. Clin. Microbiol. 35:1477-1491), automated Q-beta replicase amplification assay (see, e.g., Burg (1996) Mol. Cell. Probes 10:257-271) and other RNA polymerase mediated techniques (e.g., NASBA, Cangene, Mississauga, Ontario).

Alternatively, these nucleic acids can be synthesized *in vitro* by well-known chemical synthesis techniques, as described in, e.g., Adams (1983) J. Am. Chem. Soc. 105:661; Belousov (1997) Nucleic Acids Res. 25:3440 3444; Frenkel (1995) Free Radic. Biol. Med. 19:373 380; Blommers (1994) Biochemistry 33:7886 7896; Narang (1979) Meth. Enzymol. 68:90; Brown (1979) Meth. Enzymol. 68:109; Beaucage (1981) Tetra. Lett. 22:1859; U.S. Patent No. 4,458,066.

Techniques for the manipulation of nucleic acids, such as, e.g., subcloning, labeling probes (e.g., random-primer labeling using Klenow polymerase, nick translation, amplification), sequencing, hybridization and the like are well described in the scientific and patent literature, see, e.g., Sambrook, ed., MOLECULAR CLONING: A LABORATORY MANUAL (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Ausubel, ed. John Wiley & Sons, Inc., New York (1997); LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY: HYBRIDIZATION WITH NUCLEIC ACID PROBES, Part I. Theory and Nucleic Acid Preparation, Tijssen, ed. Elsevier, N.Y. (1993). Another useful means of obtaining and manipulating nucleic acids used to practice the methods of the invention is to clone from genomic samples, and, if desired, screen and re-clone inserts isolated or amplified from, e.g., genomic clones or cDNA clones. Sources of nucleic acid used in the methods of the invention include genomic or cDNA libraries contained in, e.g., mammalian artificial chromosomes (MACs), see, e.g., U.S. Patent Nos. 5,721,118; 6,025,155; human artificial chromosomes, see, e.g., Rosenfeld (1997) Nat. Genet. 15:333-335; yeast artificial chromosomes (YAC); bacterial artificial chromosomes (BAC); P1 artificial chromosomes, see, e.g., Woon (1998) Genomics 50:306-316; P1-derived vectors (PACs), see, e.g., Kern (1997) Biotechniques 23:120-124; cosmids, recombinant viruses, phages or plasmids.

The nucleic acids and proteins of the invention can be detected, confirmed and quantified by any of a number of means well known to those of skill in the art. General methods for detecting both nucleic acids and corresponding proteins include analytic biochemical methods such as spectrophotometry, radiography, electrophoresis, capillary electrophoresis, high performance liquid chromatography (HPLC), thin layer chromatography (TLC), hyperdiffusion chromatography, and the like, and various immunological methods such as fluid or gel precipitin reactions, immunodiffusion (single or double), immunoelectrophoresis, radioimmunoassays (RIAs), enzyme-linked immunosorbent assays (ELISAs), immunofluorescent assays, and the like. The detection of nucleic acids and polypeptides can be by well known methods such as Southern analysis, northern analysis, gel electrophoresis, PCR, radiolabeling, scintillation counting, and affinity chromatography.

In one step of an exemplary method of the invention, an esterase is used. Any esterase, or enzyme (e.g., a hydrolase) or other polypeptide having a similar activity can be used.

Capillary Arrays

The methods of the invention can be practiced in whole or in part by capillary arrays, such as the GIGAMATRIX™, Diversa Corporation, San Diego, CA. See, e.g., WO0138583. Reagents or polypeptides (e.g., enzymes) can be immobilized to or applied to an array, including capillary arrays. Capillary arrays provide another system for holding and screening reagents, catalysts (e.g., enzymes) and products. The apparatus can further include interstitial material disposed between adjacent capillaries in the array, and one or more reference indicia formed within of the interstitial material. High throughput screening apparatus can also be adapted and used to practice the methods of the invention, see, e.g., U.S. Patent Application No. 20020001809.

Whole Cell-Based Methods

The methods of the invention can be practiced in whole or in part in a whole cell environment. The invention also provides for whole cell evolution, or whole cell engineering, of a cell to develop a new cell strain having a new phenotype to be used in the methods of the invention, e.g., a new cell line comprising one, several or all enzymes used in a method of the invention. This can be done by modifying the genetic composition of the cell, where the genetic composition is modified by addition to the cell of a nucleic acid, e.g.,

a coding sequence for an enzyme used in the methods of the invention. See, e.g., WO0229032; WO0196551.

The host cell for the "whole-cell process" may be any cell known to one skilled in the art, including prokaryotic cells, eukaryotic cells, such as bacterial cells, fungal cells, yeast cells, mammalian cells, insect cells, or plant cells.

To detect the production of an intermediate or product of the methods of the invention, or a new phenotype, at least one metabolic parameter of a cell (or a genetically modified cell) is monitored in the cell in a "real time" or "on-line" time frame by Metabolic Flux Analysis (MFA). In one aspect, a plurality of cells, such as a cell culture, is monitored in "real time" or "on-line." In one aspect, a plurality of metabolic parameters is monitored in "real time" or "on-line."

Metabolic flux analysis (MFA) is based on a known biochemistry framework. A linearly independent metabolic matrix is constructed based on the law of mass conservation and on the pseudo-steady state hypothesis (PSSH) on the intracellular metabolites. In practicing the methods of the invention, metabolic networks are established, including the:

- identity of all pathway substrates, products and intermediary metabolites
- identity of all the chemical reactions interconverting the pathway metabolites, the stoichiometry of the pathway reactions,
- identity of all the enzymes catalyzing the reactions, the enzyme reaction kinetics,
- the regulatory interactions between pathway components, e.g. allosteric interactions, enzyme-enzyme interactions etc,
- intracellular compartmentalization of enzymes or any other supramolecular organization of the enzymes, and,
- the presence of any concentration gradients of metabolites, enzymes or effector molecules or diffusion barriers to their movement.

Once the metabolic network for a given strain is built, mathematic presentation by matrix notion can be introduced to estimate the intracellular metabolic fluxes if the on-line metabolome data is available. Metabolic phenotype relies on the changes of the whole metabolic network within a cell. Metabolic phenotype relies on the change of

pathway utilization with respect to environmental conditions, genetic regulation, developmental state and the genotype, etc. In one aspect of the methods of the invention, after the on-line MFA calculation, the dynamic behavior of the cells, their phenotype and other properties are analyzed by investigating the pathway utilization.

Control of physiological state of cell cultures will become possible after the pathway analysis. The methods of the invention can help determine how to manipulate the fermentation by determining how to change the substrate supply, temperature, use of inducers, etc. to control the physiological state of cells to move along desirable direction. In practicing the methods of the invention, the MFA results can also be compared with transcriptome and proteome data to design experiments and protocols for metabolic engineering or gene shuffling, etc. Any aspect of metabolism or growth can be monitored.

Monitoring expression of an mRNA transcript

In one aspect of the invention, the engineered phenotype comprises increasing or decreasing the expression of an mRNA transcript or generating new transcripts in a cell. This increased or decreased expression can be traced by use of a fluorescent polypeptide, e.g., a chimeric protein comprising an enzyme used in the methods of the invention. mRNA transcripts, or messages, also can be detected and quantified by any method known in the art, including, e.g., Northern blots, quantitative amplification reactions, hybridization to arrays, and the like. Quantitative amplification reactions include, e.g., quantitative PCR, including, e.g., quantitative reverse transcription polymerase chain reaction, or RT-PCR; quantitative real time RT-PCR, or "real-time kinetic RT-PCR" (see, e.g., Kreuzer (2001) Br. J. Haematol. 114:313-318; Xia (2001) Transplantation 72:907-914).

In one aspect of the invention, the engineered phenotype is generated by knocking out expression of a homologous gene. The gene's coding sequence or one or more transcriptional control elements can be knocked out, e.g., promoters enhancers. Thus, the expression of a transcript can be completely ablated or only decreased.

In one aspect of the invention, the engineered phenotype comprises increasing the expression of a homologous gene. This can be effected by knocking out of a negative control element, including a transcriptional regulatory element acting in cis- or trans- , or, mutagenizing a positive control element. One or more, or, all the transcripts of a cell can be measured by hybridization of a sample comprising transcripts of the cell, or, nucleic acids

representative of or complementary to transcripts of a cell, by hybridization to immobilized nucleic acids on an array.

Monitoring expression of a polypeptides, peptides and amino acids

In one aspect of the invention, the engineered phenotype comprises increasing or decreasing the expression of a polypeptide or generating new polypeptides in a cell. This increased or decreased expression can be traced by use of a fluorescent polypeptide, e.g., a chimeric protein comprising an enzyme used in the methods of the invention. Polypeptides, reagents and end products (e.g., simvastatin) also can be detected and quantified by any method known in the art, including, e.g., nuclear magnetic resonance (NMR), spectrophotometry, radiography (protein radiolabeling), electrophoresis, capillary electrophoresis, high performance liquid chromatography (HPLC), thin layer chromatography (TLC), hyperdiffusion chromatography, various immunological methods, e.g. immunoprecipitation, immunodiffusion, immuno-electrophoresis, radioimmunoassays (RIAs), enzyme-linked immunosorbent assays (ELISAs), immuno-fluorescent assays, gel electrophoresis (e.g., SDS-PAGE), staining with antibodies, fluorescent activated cell sorter (FACS), pyrolysis mass spectrometry, Fourier-Transform Infrared Spectrometry, Raman spectrometry, GC-MS, and LC-Electrospray and cap-LC-tandem-electrospray mass spectrometries, and the like. Novel bioactivities can also be screened using methods, or variations thereof, described in U.S. Patent No. 6,057,103. Polypeptides of a cell can be measured using a protein array.

Determining the degree of sequence identity

In one aspect of any of the methods of the invention, at least one enzymatic reaction is carried out by a hydrolase (e.g., an esterase, or acylase) encoded by a nucleic acid having at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or complete (100%) sequence identity to SEQ ID NO:1, SEQ ID NO:3 and/or SEQ ID NO:5, or enzymatically active fragments thereof. In one aspect of any of the methods of the invention, at least one enzymatic reaction is carried out by a hydrolase (e.g., an esterase, or acylase) having a sequence at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%,

68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or complete (100%) sequence identity to SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6, or enzymatically active fragments thereof..

Enzymatic activity can be determined by routine screening using known protocols, or, the methods of the invention, as described herein. For example, enzymatic activity can be determined by testing whether a polypeptide or peptide can hydrolyze a lactone ring, or, enzymatically acylate a diol lactone, as described herein.

Protein and/or nucleic acid sequence homologies may be evaluated using any of the variety of sequence comparison algorithms and programs known in the art. Such algorithms and programs include, but are by no means limited to, TBLASTN, BLASTP, FASTA, TFASTA, and CLUSTALW (see, e.g., Pearson (1988) Proc. Natl. Acad. Sci. USA 85(8):2444-2448; Altschul (1990) J. Mol. Biol. 215(3):403-410; Thompson (1994) Nucleic Acids Res. 22(2):4673-4680; Higgins *et al.*, Methods Enzymol. 266:383-402, 1996; Altschul *et al.*, J. Mol. Biol. 215(3):403-410, 1990; Altschul *et al.*, Nature Genetics 3:266-272, 1993).

Homology or identity is often measured using sequence analysis software (*e.g.*, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705). Such software matches similar sequences by assigning degrees of homology to various deletions, substitutions and other modifications. The terms "homology" and "identity" in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same when compared and aligned for maximum correspondence over a comparison window or designated region as measured using any number of sequence comparison algorithms or by manual alignment and visual inspection.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The

sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

A "comparison window", as used herein, includes reference to a segment of any one of the numbers of contiguous residues. For example, in alternative aspects of the invention, contiguous residues ranging anywhere from about 20 to the full length of an exemplary polypeptide or nucleic acid sequence of the invention are compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. If the reference sequence has the requisite sequence identity to an exemplary polypeptide or nucleic acid sequence of the invention, e.g., 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5 or SEQ ID NO:6, and the sequence is or encodes a hydrolase, that sequence can be used in at least one step of a method of the invention. In alternative embodiments, subsequences ranging from about 20 to 600, about 50 to 200, and about 100 to 150 are compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequence for comparison are well known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, Adv. Appl. Math. 2:482, 1981, by the homology alignment algorithm of Needleman & Wunsch, J. Mol. Biol. 48:443, 1970, by the search for similarity method of person & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444, 1988, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection. Other algorithms for determining homology or identity include, for example, in addition to a BLAST program (Basic Local Alignment Search Tool at the National Center for Biological Information), ALIGN, AMAS (Analysis of Multiply Aligned Sequences), AMPS (Protein Multiple Sequence Alignment), ASSET (Aligned Segment Statistical Evaluation Tool), BANDS, BESTSCOR, BIOSCAN (Biological Sequence Comparative Analysis Node), BLIMPS (BLocks IMProved Searcher), FASTA, Intervals & Points, BMB, CLUSTAL V, CLUSTAL

W, CONSENSUS, LCONSENSUS, WCONSENSUS, Smith-Waterman algorithm, DARWIN, Las Vegas algorithm, FNAT (Forced Nucleotide Alignment Tool), Framealign, Framesearch, DYNAMIC, FILTER, FSAP (Fristensky Sequence Analysis Package), GAP (Global Alignment Program), GENAL, GIBBS, GenQuest, ISSC (Sensitive Sequence Comparison), LALIGN (Local Sequence Alignment), LCP (Local Content Program), MACAW (Multiple Alignment Construction & Analysis Workbench), MAP (Multiple Alignment Program), MBLKP, MBLKN, PIMA (Pattern-Induced Multi-sequence Alignment), SAGA (Sequence Alignment by Genetic Algorithm) and WHAT-IF. Such alignment programs can also be used to screen genome databases to identify polynucleotide sequences having substantially identical sequences. Databases containing genomic information annotated with some functional information are maintained by different organization, and are accessible via the internet.

BLAST, BLAST 2.0 and BLAST 2.2.2 algorithms are also used to practice the invention. They are described, e.g., in Altschul (1977) Nuc. Acids Res. 25:3389-3402; Altschul (1990) J. Mol. Biol. 215:403-410. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul (1990) supra). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4 and

a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectations (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915) alignments (B) of 50, expectation (E) of 10, M=5, N= -4, and a comparison of both strands. The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873). One measure of similarity provided by BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a references sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001. In one aspect, protein and nucleic acid sequence homologies are evaluated using the Basic Local Alignment Search Tool ("BLAST"). For example, five specific BLAST programs can be used to perform the following task: (1) BLASTP and BLAST3 compare an amino acid query sequence against a protein sequence database; (2) BLASTN compares a nucleotide query sequence against a nucleotide sequence database; (3) BLASTX compares the six-frame conceptual translation products of a query nucleotide sequence (both strands) against a protein sequence database; (4) TBLASTN compares a query protein sequence against a nucleotide sequence database translated in all six reading frames (both strands); and, (5) TBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database. The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are preferably identified (i.e., aligned) by means of a scoring matrix, many of which are known in the art. Preferably, the scoring matrix used is the BLOSUM62 matrix (Gonnet et al., Science 256:1443-1445, 1992; Henikoff and Henikoff, Proteins 17:49-61, 1993). Less preferably, the PAM or PAM250 matrices may also be used (see, e.g., Schwartz and Dayhoff, eds., 1978, Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure, Washington: National Biomedical Research Foundation).

In one aspect of the invention, the NCBI BLAST 2.2.2 programs is used, default options to blastp. There are about 38 setting options in the BLAST 2.2.2 program. In this exemplary aspect of the invention, all default values are used except for the default filtering setting (i.e., all parameters set to default except filtering which is set to OFF); in its place a "-F F" setting is used, which disables filtering. Use of default filtering often results in Karlin-Altschul violations due to short length of sequence.

The default values used in this exemplary aspect of the invention include:

"Filter for low complexity: ON
Word Size: 3
Matrix: Blosum62
Gap Costs: Existence:11
Extension:1"

Other default settings can be: filter for low complexity OFF, word size of 3 for protein, BLOSUM62 matrix, gap existence penalty of -11 and a gap extension penalty of -1. An exemplary NCBI BLAST 2.2.2 program setting has the "-W" option default to 0. This means that, if not set, the word size defaults to 3 for proteins and 11 for nucleotides.

The invention will be further described with reference to the following examples; however, it is to be understood that the invention is not limited to such examples.

EXAMPLES

Example 1: Chemoenzymatic production of Simvastatin

The following example describes an exemplary protocol of the invention for the chemoenzymatic production of Simvastatin.

Enzymatic Hydrolysis of Lovastatin (Step 1, above)

The enzyme having a sequence as set forth in SEQ ID NO:2 (encoded by SEQ ID NO:1) was evaluated at 0.1 to 0.5 M concentrations of lovastatin or lovastatin acid in 7-10% MeOH/buffer, with the reaction being maintained at pH 9-9.5 by automatic addition of base. The best result was obtained at 0.5M lovastatin on a 500 mL scale using a lyophilized preparation of enzyme SEQ ID NO:2 (centrifuged supernatant from lysed cells) containing 14 mg/mL total protein; complete conversion of substrate was observed after 48 h.

The reaction mixture was acidified (pH 2), and the precipitate collected by centrifugation and dried. The filtrate was extracted with iPrOAc and the organic extract was added to the dried filter cake. The resulting suspension was heat to reflux in a Dean-Stark apparatus until lactonization was complete. The resulting solution was filtered through a Celite pad, and the filtrate was washed with satd. $NaHCO_3$. The resulting iPrOAc solution was concentrated until (x 0.5), diluted with hexanes and cooled to 0°C. The precipitated solid was filtered and air-dried to yield diol lactone (63 g, 79.5% isolated yield; another 10.3 g of product was identified in various washes and mother liquors). The product contained <1% lovastatin.

Enzymatic Acylation of Diol Lactone (Step 3, above)

A mixture of diol lactone (25 mM), vinyl acetate (250 mM) and *Candida antarctica* lipase B (33 mg) in TBME (1 mL) was shaken at RT. After 44 h HPLC indicated the formation of the monoacetate with 60% conversion.

Acetylation of Diol Lactone (Step 3, above)

Diol lactone (10 g, 31.25 mmol,) and DMAP (0.5727 g, 4.69 mmol, 15 mol%) were dissolved in anhydrous CH_2Cl_2 (62.5 ml), stirred under N_2 and cooled to 0°C. The pivalic/acetic mixed anhydride (4.95, 34.4 mmol, 1.1 equivalent) was added in two portions. The first portion (2 ml) was added in one shot, followed by the rest of anhydride added by syringe pump over 20 min. The reaction mixture was stirred 0°C for 30min, at ambient temperature for 1.5 hours. The reaction was quenched by adding 31.2ml water and stirred for 10min at ambient temperature, then the mixture was transferred into a separation funnel and the organic layer was washed sequentially by 5% HCl (31.3 ml), saturated $NaHCO_3$ (32 ml) and brine (32 ml). The organic layer was collected and dried over Na_2SO_4, concentrated after the removal of drying agent by filtration. The residue was dried *in vacuo* overnight. A slightly yellowish solid was obtained (10.8g, yield 95.5%). HPLC analysis indicated the following distribution of products: 4-acetyl lactone (95.3%), diol lactone (2.1%), 4, 8'-diacetyl lactone (1.2%), elimination (1.4%).

Preparation of Acetyl Simvastatin (Step 4, above)

4-Acetyl lactone was dried under vacuum overnight at room temperature, stored under nitrogen, then dissolved in anhydrous methylene chloride (1g/2.5-3ml ratio) at

room temperature under nitrogen. Meanwhile, $Cu(OTf)_2$ (5mol%) was dissolved in the minimum amount of acetonitrile at room temperature, then 1.05-1.2 eq of dimethylbutyric anhydride was added to the solution, stirring at room temperature for 30 min to hour. This $Cu(OTf)_2$/anhydride solution was transferred into the 4-Acetyl lactone solution through syringe at room temperature under nitrogen with stirring. When complete (monitored by HPLC), the reaction was quenched by addition of water, and washed with satd., $NaHCO_3$ The isolated organic layer was dried over Na_2SO_4, filtered and evaporated to obtain crude 4-acetyl simvastatin (>99%).

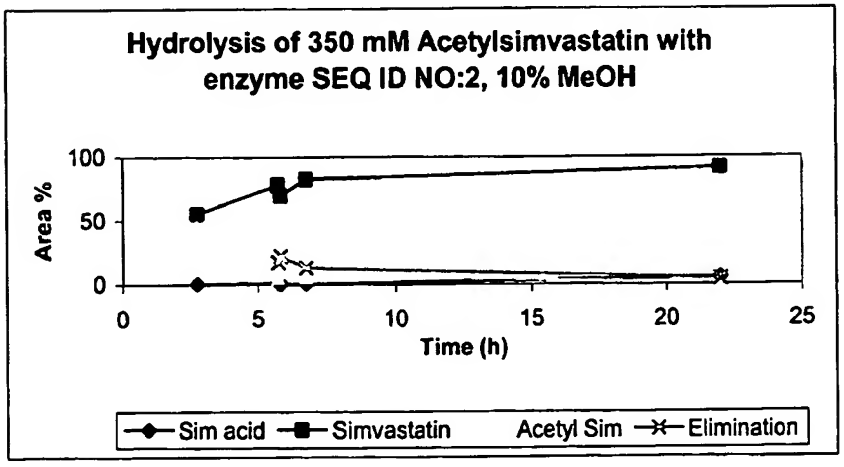
Enzymatic Hydrolysis of Acetyl Simvastatin (Step 5, above)

3.22 g Acetylsimvastatin (final concentration 350 mM)

2 ml MeOH; 100 µl 4M Tris; 9.9 ml water

8 ml BD12785 (125 mg/ml lyophilized lysate in water)

The reaction is performed in a 25 ml vessel with overhead stirring and a magnetic stirrer bar. pH-stat conditions are maintained by a DasGip STIRRER-PRO® system; a pH of 7 is maintained by addition of 10% NH_4OH. As the conversion approaches ~75%, 4 ml of toluene are added to solubilize the material. The reaction is allowed to proceed overnight, at which time further solvent (toluene or methylene chloride) is added to ensure that all insoluble material is dissolved. A sample is analyzed by HPLC.



Final composition of the reaction: Simvastatin acid 4.7%, Simvastatin 90.9%, Acetyl simvastatin 0.9%, Putative elimination product of simvastatin 3.5%. Final conversion 95.6%.

A number of embodiments of the invention have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the invention. Accordingly, other embodiments are within the scope of the following claims.

WHAT IS CLAIMED IS:

1. A method for the preparation of simvastatin comprising (i) at least one protocol as set forth in Appendix A or Appendix B, or, (ii) a process comprising steps 1 through 5, wherein

step 1 comprises enzymatic hydrolysis of lovastatin, lovastatin acid or a salt of lovastatin acid to form the triol acid;

step 2 comprises heating the triol acid or stirring in the presence of acid to form the diol lactone;

step 3 comprise protection of the 4'-OH on the lactone ring by regioselective acylation either chemically or enzymatically;

step 4 comprises acylation of the hydroxyl at the 8-position carried out chemically or enzymatically using a hydrolase; and

step 5 comprises selective removal of the acyl protecting group at the 4' position either chemically or enzymatically, thereby yielding simvastatin.

2. The method of claim 1, wherein at least one step is performed in a reaction vessel.

3. The method of claim 1, wherein at least one step is performed in a cell extract.

4. The method of claim 1, wherein at least one step is performed in a whole cell.

5. The method of claim 1, wherein an ammonium salt of simvastatin is formed.

6. The method of claim 1, further comprising crystallization of the simvastatin.

7. The method of claim 6, further comprising re-crystallization of the simvastatin.

8. The method of claim 6, further comprising relactonization to provide simvastatin with a desired purity.

9. The method of claim 1, wherein at least one enzymatic reaction is carried out by a hydrolase encoded by a nucleic acid having at least 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or complete (100%) sequence identity to SEQ ID NO:1, or enzymatically active fragments thereof..

10. The method of claim 1, wherein at least one enzymatic reaction is carried out by a hydrolase encoded by a nucleic acid having at least 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or complete (100%) sequence identity to SEQ ID NO:3, or enzymatically active fragments thereof..

11. The method of claim 1, wherein at least one enzymatic reaction is carried out by a hydrolase encoded by a nucleic acid having at least 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or complete (100%) sequence identity to SEQ ID NO:5, or enzymatically active fragments thereof..

12. The method of claim 1, wherein at least one enzymatic reaction is carried out by a hydrolase having a sequence at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or complete (100%) sequence identity to SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6, or enzymatically active fragments thereof.

ABSTRACT

METHODS FOR MAKING SIMVASTATIN AND INTERMEDIATES

The invention provides synthetic chemical and chemoenzymatic methods of producing simvastatin and various intermediates. In one aspect, enzymes such as hydrolases are used in the methods of the invention.

SEQUENCE LISTING

<110> Burk, Mark
Zhu, Zoulin
Chaplin, Jennifer
Kustedjo, Karen
Huang, Zilin
Morgan , Brian
Levin, Michael

<120> METHODS FOR MAKING SIMVASTATIN AND INTERMEDIATES

<130> 09010-983P01

<140> not assigned
<141> 2003-10-21

<160> 6

<170> PatentIn version 3.1

<210> 1
<211> 1629
<212> DNA
<213> Unknown

<220>
<223> Environmental

<400> 1
atgagccttt gcgtcattcg attcatcgcc ggaactttgg tacttgtggc gtcagtggaa 60

tcggcagttg ctcaacaagc gtgtgctgac ctgatgggcc tcgagctgcc gtatacaacg 120

ataacgtccg ctgcagtggc taccgagggc ccaatcccac agccggcgat ctttggaagc 180

actgacccca ttgtggctcc agagcgatgt gaagtgcggg cggtcacgcg ccctacgaag 240

gactccgaga ttcgaatcga gctctggctg ccgctctccg gatggaacgg aaagtatcta 300

caaattggta gcggtggctg ggctggttcg atcaatcgaa cggggctgat aggccctctt 360

cagcgcggtt atgccgtagc cgcaaccgac aatggccata tcagcgaagg tttggtgcct 420

gacgcctcct gggctatcgg ccatccgcaa aagctgatcg atttcggtta tcgcgccgtg 480

cacgaaacaa gtgttcaggc caaagctatc ctgcgcgcct actttggccg cggtcaggat 540

ctgagctact tcagcggttg ttctaatggc ggacgcgagg ctctcatgga ggcgcagcgc 600

tatccggaag atttcgaagg catcatcgcc ggagcgcccg cgaacaattg gtcgcgcctg 660

tttacggggt ttgtgtggaa tgaacgcgcg ttggcggacg atccaattcc tcctgccaag 720

ttgacagcga ttcaggcggc ggcaattgct gcgtgtgata cgctggacgg tgttgaggac 780

gggctcatcg aaaacccacg agcgtgtagc ttcgatccgc gttcaatggt ctgtacagcc 840

gatgatgcct ctgactgtct gacagaagga caggtcgcga cgctacacag gatatatagc 900

ggcccaacca atcctcggac cggtgagcga atctttccag gctatccgat gggcaccgaa 960

gccgtgccgg gcggatgggt accgtggatc gtgtccgcga gctccgaagt tccgagcata 1020

caagcaagct ttggcaactc ctattacggg cacgcggtct tcgagcaatc gaactgggat 1080

ttcaggacgt tggatttcga ccaggacgtt gcgtttggcg atgcgaaggc ggggccggtg 1140

ctcaatgcca cgaaccccga tctgcgttcg tttcgcgcga atggcggcaa actgattcag 1200

tatcatggct ggggcgatgc agccattacg gcttttagtt cgatcgacta ctacgagaac 1260

gtgcgcgcct tcctcgatcg cttccccgac ccccgaagcg agaacacgga tatcgacggt 1320

ttctatcgcc tgttcctggt tccgggcatg ggacattgct ccggcgggat cggcccaagt 1380

agctttggca atggcttccg ttccgcacgt acggatgccg agcacgacct actctccgcc 1440

cttgaggcat gggtggagcg agacacggcc ccggagagat tgatcggaac ggggacggcc 1500

gtaggcgacc caaccgcgac tctgacgcgt ccgctatgcc catatccgcg gacggcacgg 1560

tatctcggaa gcggcaactc aaatgatgcg gccaacttcg agtgcgccct gcccgctggc 1620

gtgcagtag 1629

<210> 2
<211> 542
<212> PRT
<213> Unkown

<220>
<223> Environmental

<220>
<221> SIGNAL
<222> (1)...(24)

<400> 2
Met Ser Leu Cys Val Ile Arg Phe Ile Ala Gly Thr Leu Val Leu Val
1 5 10 15

Ala Ser Val Glu Ser Ala Val Ala Gln Gln Ala Cys Ala Asp Leu Met
20 25 30

Gly Leu Glu Leu Pro Tyr Thr Thr Ile Thr Ser Ala Ala Val Ala Thr
35 40 45

Glu Gly Pro Ile Pro Gln Pro Ala Ile Phe Gly Ser Thr Asp Pro Ile
50 55 60

```
Val Ala Pro Glu Arg Cys Glu Val Arg Ala Val Thr Arg Pro Thr Lys
65                  70                  75                  80

Asp Ser Glu Ile Arg Ile Glu Leu Trp Leu Pro Leu Ser Gly Trp Asn
                85                  90                  95

Gly Lys Tyr Leu Gln Ile Gly Ser Gly Gly Trp Ala Gly Ser Ile Asn
            100                 105                 110

Arg Thr Gly Leu Ile Gly Pro Leu Gln Arg Gly Tyr Ala Val Ala Ala
        115                 120                 125

Thr Asp Asn Gly His Ile Ser Glu Gly Leu Val Pro Asp Ala Ser Trp
    130                 135                 140

Ala Ile Gly His Pro Gln Lys Leu Ile Asp Phe Gly Tyr Arg Ala Val
145                 150                 155                 160

His Glu Thr Ser Val Gln Ala Lys Ala Ile Leu Arg Ala Tyr Phe Gly
                165                 170                 175

Arg Gly Gln Asp Leu Ser Tyr Phe Ser Gly Cys Ser Asn Gly Gly Arg
            180                 185                 190

Glu Ala Leu Met Glu Ala Gln Arg Tyr Pro Glu Asp Phe Glu Gly Ile
        195                 200                 205

Ile Ala Gly Ala Pro Ala Asn Asn Trp Ser Arg Leu Phe Thr Gly Phe
    210                 215                 220

Val Trp Asn Glu Arg Ala Leu Ala Asp Asp Pro Ile Pro Pro Ala Lys
225                 230                 235                 240

Leu Thr Ala Ile Gln Ala Ala Ala Ile Ala Ala Cys Asp Thr Leu Asp
                245                 250                 255

Gly Val Glu Asp Gly Leu Ile Glu Asn Pro Arg Ala Cys Ser Phe Asp
            260                 265                 270

Pro Arg Ser Met Val Cys Thr Ala Asp Asp Ala Ser Asp Cys Leu Thr
        275                 280                 285
```

```
Glu Gly Gln Val Ala Thr Leu His Arg Ile Tyr Ser Gly Pro Thr Asn
    290                 295                 300

Pro Arg Thr Gly Glu Arg Ile Phe Pro Gly Tyr Pro Met Gly Thr Glu
305                 310                 315                 320

Ala Val Pro Gly Gly Trp Val Pro Trp Ile Val Ser Ala Ser Ser Glu
                325                 330                 335

Val Pro Ser Ile Gln Ala Ser Phe Gly Asn Ser Tyr Tyr Gly His Ala
            340                 345                 350

Val Phe Glu Gln Ser Asn Trp Asp Phe Arg Thr Leu Asp Phe Asp Gln
        355                 360                 365

Asp Val Ala Phe Gly Asp Ala Lys Ala Gly Pro Val Leu Asn Ala Thr
    370                 375                 380

Asn Pro Asp Leu Arg Ser Phe Arg Ala Asn Gly Gly Lys Leu Ile Gln
385                 390                 395                 400

Tyr His Gly Trp Gly Asp Ala Ala Ile Thr Ala Phe Ser Ser Ile Asp
                405                 410                 415

Tyr Tyr Glu Asn Val Arg Ala Phe Leu Asp Arg Phe Pro Asp Pro Arg
            420                 425                 430

Ser Glu Asn Thr Asp Ile Asp Gly Phe Tyr Arg Leu Phe Leu Val Pro
        435                 440                 445

Gly Met Gly His Cys Ser Gly Gly Ile Gly Pro Ser Ser Phe Gly Asn
    450                 455                 460

Gly Phe Arg Ser Ala Arg Thr Asp Ala Glu His Asp Leu Leu Ser Ala
465                 470                 475                 480

Leu Glu Ala Trp Val Glu Arg Asp Thr Ala Pro Glu Arg Leu Ile Gly
                485                 490                 495

Thr Gly Thr Ala Val Gly Asp Pro Thr Ala Thr Leu Thr Arg Pro Leu
            500                 505                 510
```

Cys Pro Tyr Pro Arg Thr Ala Arg Tyr Leu Gly Ser Gly Asn Ser Asn
515 520 525

Asp Ala Ala Asn Phe Glu Cys Ala Leu Pro Ala Gly Val Gln
530 535 540

<210> 3
<211> 1209
<212> DNA
<213> Unknown

<220>
<223> Environmental

<400> 3
atggaaatcc atggtacatg cgacccaaag tttcacttgg tgcggcagga gtttgaacga 60

aatttgcgtg agcgcggcga agtaggagcg tccgtttgcg tcacgttgca cggcgaaacc 120

gtagtggact tgtggggcgg catggcgcgt gccgacactc agacgccatg gacggcggag 180

acggtcagta ttgttttttc ctccaccaaa ggcgcaacgg cactctgcgc ccatatgctg 240

gcgtcacgcg gccaactgga tcttgatgca ccagtcgcca cctactggcc ggaatttgcc 300

caagccggca aagctcgcat cccggtgaaa atgctcttga accatcaagc tggtctccct 360

gccgtacgga caccgctgcc ccagggtgcc tacgctgact gggaactgat ggtcaatacg 420

ttggccaagg aagagccgtt ttgggaacct ggcacccgca acggctatca tgcgctcacc 480

atggggtggc tggtgggaga agtggtgcga cgtgtctctg gtaagtcgct tgggacattc 540

ttccaagagg agatcgccag gccgttgggg ttagatttct ggattggctt accagcagag 600

caagaggcac gggtcgcgcc gatgatcgcg gcggagcctg atccgcaaag cctcttcttc 660

caagaggtcg cgaagcctgg ggccttacag tcgctcgtac tccttaactc cggcggctat 720

atgggtgctc agcctgagta tgactcgcgg gcggcgcatg cggccgagat tggtgcagcc 780

ggtggtatca ccaacgcacg cggcctggca ggcatgtacg caccactggc ctgcggaggc 840

aaactcaaag gggtggagtt ggtcagtcct gacatgctgg cccgaatgtc cagagtggcc 900

tctgcgactg ggagagatgc cgtgctcatg atgccaaccc ggtttgccct gggcttcatg 960

aagtccatgg acaaccgccg ggagcctgct ggcgtgcagg acagcgcgct ctttggggag 1020

gaggcttttg gccatgtggg ggccgggggt tcgtttggtt ttgccgatcc caaagcagga 1080

atgtcctttg gctataccat gaaccgaatg gggctgggag ccgggctcaa cccgcggggg 1140

caaagcctgg tggatgcaac ctaccgctcg ttagggtatc agtcggatgc ctctggagcc 1200

tggacctga 1209

<210> 4
<211> 402
<212> PRT
<213> Unkown

<220>
<223> Environmental

<220>
<221> DOMAIN
<222> (18)...(386)
<223> Beta-lactamase

<400> 4
Met Glu Ile His Gly Thr Cys Asp Pro Lys Phe His Leu Val Arg Gln
1 5 10 15

Glu Phe Glu Arg Asn Leu Arg Glu Arg Gly Glu Val Gly Ala Ser Val
20 25 30

Cys Val Thr Leu His Gly Glu Thr Val Val Asp Leu Trp Gly Gly Met
35 40 45

Ala Arg Ala Asp Thr Gln Thr Pro Trp Thr Ala Glu Thr Val Ser Ile
50 55 60

Val Phe Ser Ser Thr Lys Gly Ala Thr Ala Leu Cys Ala His Met Leu
65 70 75 80

Ala Ser Arg Gly Gln Leu Asp Leu Asp Ala Pro Val Ala Thr Tyr Trp
85 90 95

Pro Glu Phe Ala Gln Ala Gly Lys Ala Arg Ile Pro Val Lys Met Leu
100 105 110

Leu Asn His Gln Ala Gly Leu Pro Ala Val Arg Thr Pro Leu Pro Gln
115 120 125

Gly Ala Tyr Ala Asp Trp Glu Leu Met Val Asn Thr Leu Ala Lys Glu
130 135 140

Glu Pro Phe Trp Glu Pro Gly Thr Arg Asn Gly Tyr His Ala Leu Thr
145 150 155 160

Met Gly Trp Leu Val Gly Glu Val Val Arg Arg Val Ser Gly Lys Ser
165 170 175

Leu Gly Thr Phe Phe Gln Glu Glu Ile Ala Arg Pro Leu Gly Leu Asp
180 185 190

Phe Trp Ile Gly Leu Pro Ala Glu Gln Glu Ala Arg Val Ala Pro Met
195 200 205

Ile Ala Ala Glu Pro Asp Pro Gln Ser Leu Phe Phe Gln Glu Val Ala
210 215 220

Lys Pro Gly Ala Leu Gln Ser Leu Val Leu Leu Asn Ser Gly Gly Tyr
225 230 235 240

Met Gly Ala Gln Pro Glu Tyr Asp Ser Arg Ala Ala His Ala Ala Glu
245 250 255

Ile Gly Ala Ala Gly Gly Ile Thr Asn Ala Arg Gly Leu Ala Gly Met
260 265 270

Tyr Ala Pro Leu Ala Cys Gly Gly Lys Leu Lys Gly Val Glu Leu Val
275 280 285

Ser Pro Asp Met Leu Ala Arg Met Ser Arg Val Ala Ser Ala Thr Gly
290 295 300

Arg Asp Ala Val Leu Met Met Pro Thr Arg Phe Ala Leu Gly Phe Met
305 310 315 320

Lys Ser Met Asp Asn Arg Arg Glu Pro Ala Gly Val Gln Asp Ser Ala
325 330 335

Leu Phe Gly Glu Glu Ala Phe Gly His Val Gly Ala Gly Gly Ser Phe
340 345 350

Gly Phe Ala Asp Pro Lys Ala Gly Met Ser Phe Gly Tyr Thr Met Asn
355 360 365

Arg Met Gly Leu Gly Ala Gly Leu Asn Pro Arg Gly Gln Ser Leu Val
370 375 380

Asp Ala Thr Tyr Arg Ser Leu Gly Tyr Gln Ser Asp Ala Ser Gly Ala

385 390 395 400

Trp Thr

<210> 5
<211> 1578
<212> DNA
<213> Unknown

<220>
<223> Environmental

<400> 5
atgagatcag cagctcgcat cagcgtggcg gcagttgcct ttctttgcct gctcttgacg 60

actcgggttt ccgcccagat cgtgccggcg atggaatgtg cggatctggc gaatcagcag 120

cttcccaaca cgacgatcac ctcggcccag accgtcacca ccggatcgtt aacgcccccg 180

ggctcgacga atccgatcac cgacctgcct cctttctgcc gtgtcacagg cgccatcgcc 240

ccgacgagcg agtcgcacat cctcttcgag gtctggctgc cgctggataa atggaacggc 300

aagttcgccg gcgtgggcaa cggcggctgg gccggcatca tctccttcgg cgccctcgga 360

agccagctca agcgcggcta cgcgaccgcc tccacgaata cgggtcacga agcggcgccg 420

gggatgaacg cagccaggtt tgcgttcgag aagccggagc agcttatcga cttcgcctat 480

cgctcccagc acgagacggc cctgaaagcg aaggcgctgg ttcaggcttt ctacgggaag 540

ccgccggaac actcctattt catcgggtgc tcatcgggtg ggtaccaggg cctgatggag 600

gcccaacgat ttccggccga ctacgacggg atcgtcgccg gtatgccggc gaacaactgg 660

acacggctga tggccggcga cttggacgcg atccttgccg tctccgtaga tcctgccagc 720

caccttcccg tctccgcatt gggtctgttg tatcgctcgg tgctcgctgc ctgcgacggc 780

atcgacggtg ttgtagacgg tgttctggag gatccgcgcc gatgccggtt cgacccggcc 840

gtgttgatgt gcaaggcgga tcagaatccc gatggctgcc ttacgccggc tcaggtggaa 900

gcggcacggc gcatatacgg cggtctgaag gatcccaaga ccggcgctca gctctatccg 960

gggctggcgc cgggaagcga gccgttctgg ccgcaccgca atccggcgaa tccgttccct 1020

attccgatcg cgcactacaa gtggctcgtc tttgccgatc caaactggga ttggagaaca 1080

ttcaagttca cggatccggc ggactaccag gctttcctca atgcggaagc cacgtatgcc 1140

cctactctca atgcgaccaa tccggacctc cgggagttca gccggcgcgg cggcaggttg 1200

attcagtacc atggctggaa cgatcagctg attgccccgc aaaacagcat cgactattac 1260

gagagcgtcc tttcgttctt cgggtccggc aaacaggatc gagcgcagac cgtgcgcgag 1320

gttcagagct tctaccggct gttcatggcg ccgggtatgg ctcactgtgg aggcggtaca 1380

ggtccgaact catttgacat gctggatgcc ctcgagaagt gggtggaagg cgggatagcg 1440

ccggaacgag tccttgcgac gcgttccata aacggcgtag tcgaccggct gcgcccgctc 1500

tgtccatatc cgcaggtcgc cgtgtacaag ggtcatgggg atacaaacga cgccgcgaac 1560

ttcgtctgtc gcgattag 1578

<210> 6
<211> 525
<212> PRT
<213> Unkown

<220>
<223> Environmental

<220>
<221> SIGNAL
<222> (1)...(25)

<400> 6
Met Arg Ser Ala Ala Arg Ile Ser Val Ala Ala Val Ala Phe Leu Cys
1 5 10 15

Leu Leu Leu Thr Thr Arg Val Ser Ala Gln Ile Val Pro Ala Met Glu
20 25 30

Cys Ala Asp Leu Ala Asn Gln Gln Leu Pro Asn Thr Thr Ile Thr Ser
35 40 45

Ala Gln Thr Val Thr Thr Gly Ser Leu Thr Pro Pro Gly Ser Thr Asn
50 55 60

Pro Ile Thr Asp Leu Pro Pro Phe Cys Arg Val Thr Gly Ala Ile Ala
65 70 75 80

Pro Thr Ser Glu Ser His Ile Leu Phe Glu Val Trp Leu Pro Leu Asp
85 90 95

Lys Trp Asn Gly Lys Phe Ala Gly Val Gly Asn Gly Gly Trp Ala Gly
100 105 110

Ile Ile Ser Phe Gly Ala Leu Gly Ser Gln Leu Lys Arg Gly Tyr Ala
115 120 125

```
Thr Ala Ser Thr Asn Thr Gly His Glu Ala Ala Pro Gly Met Asn Ala
    130                 135                 140

Ala Arg Phe Ala Phe Glu Lys Pro Glu Gln Leu Ile Asp Phe Ala Tyr
145                 150                 155                 160

Arg Ser Gln His Glu Thr Ala Leu Lys Ala Lys Ala Leu Val Gln Ala
                165                 170                 175

Phe Tyr Gly Lys Pro Pro Glu His Ser Tyr Phe Ile Gly Cys Ser Ser
            180                 185                 190

Gly Gly Tyr Gln Gly Leu Met Glu Ala Gln Arg Phe Pro Ala Asp Tyr
        195                 200                 205

Asp Gly Ile Val Ala Gly Met Pro Ala Asn Asn Trp Thr Arg Leu Met
    210                 215                 220

Ala Gly Asp Leu Asp Ala Ile Leu Ala Val Ser Val Asp Pro Ala Ser
225                 230                 235                 240

His Leu Pro Val Ser Ala Leu Gly Leu Leu Tyr Arg Ser Val Leu Ala
                245                 250                 255

Ala Cys Asp Gly Ile Asp Gly Val Val Asp Gly Val Leu Glu Asp Pro
            260                 265                 270

Arg Arg Cys Arg Phe Asp Pro Ala Val Leu Met Cys Lys Ala Asp Gln
        275                 280                 285

Asn Pro Asp Gly Cys Leu Thr Pro Ala Gln Val Glu Ala Ala Arg Arg
    290                 295                 300

Ile Tyr Gly Gly Leu Lys Asp Pro Lys Thr Gly Ala Gln Leu Tyr Pro
305                 310                 315                 320

Gly Leu Ala Pro Gly Ser Glu Pro Phe Trp Pro His Arg Asn Pro Ala
                325                 330                 335

Asn Pro Phe Pro Ile Pro Ile Ala His Tyr Lys Trp Leu Val Phe Ala
            340                 345                 350
```

```
Asp Pro Asn Trp Asp Trp Arg Thr Phe Lys Phe Thr Asp Pro Ala Asp
        355                 360                 365

Tyr Gln Ala Phe Leu Asn Ala Glu Ala Thr Tyr Ala Pro Thr Leu Asn
    370                 375                 380

Ala Thr Asn Pro Asp Leu Arg Glu Phe Ser Arg Arg Gly Gly Arg Leu
385                 390                 395                 400

Ile Gln Tyr His Gly Trp Asn Asp Gln Leu Ile Ala Pro Gln Asn Ser
                405                 410                 415

Ile Asp Tyr Tyr Glu Ser Val Leu Ser Phe Phe Gly Ser Gly Lys Gln
            420                 425                 430

Asp Arg Ala Gln Thr Val Arg Glu Val Gln Ser Phe Tyr Arg Leu Phe
        435                 440                 445

Met Ala Pro Gly Met Ala His Cys Gly Gly Gly Thr Gly Pro Asn Ser
    450                 455                 460

Phe Asp Met Leu Asp Ala Leu Glu Lys Trp Val Glu Gly Gly Ile Ala
465                 470                 475                 480

Pro Glu Arg Val Leu Ala Thr Arg Ser Ile Asn Gly Val Val Asp Arg
                485                 490                 495

Leu Arg Pro Leu Cys Pro Tyr Pro Gln Val Ala Val Tyr Lys Gly His
            500                 505                 510

Gly Asp Thr Asn Asp Ala Ala Asn Phe Val Cys Arg Asp
        515                 520                 525
```

Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US04/034913

International filing date: 20 October 2004 (20.10.2004)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/513,237
Filing date: 21 October 2003 (21.10.2003)

Date of receipt at the International Bureau: 17 December 2004 (17.12.2004)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse